CHAPTER 6

Discussion

CHAPTER 6

Discussion

6.1 Introduction

The study of the structure-function relationship of ArgR is particularly intriguing because of the uncommon features of ArgR, which functions as a hexamer in at least two disparate biological processes. Firstly, in gene regulation, where ArgR represses expression of eleven L-arginine biosynthetic genes as well as its own synthesis by interacting with eight promoters (Glandsdorff, 1996). Secondly, ArgR also functions as an accessory factor in Xer-site specific recombination at *cer* on the multicopy plasmid ColE1 and its natural derivative (Stirling *et al.*, 1988; Burket *et al.*, 1994; Chen *et al.*, 1997).

In this study, two N-terminal fusion derivatives of ArgRWT and ArgRNV were successfully constructed and expressed in *E. coli* K-12. Two recombinant plasmids were constructed in this study that directs the synthesis of proteins as fusion of ArgR and biotinylated peptide. The *argRWT* (plasmid pAM204) and *argRNV* genes (plasmid pAM401) (Merican, 1995) were fused separately to the C-terminal of a DNA segment that codes for a biotinylated peptide present on plasmid PinPoint™ Xa-3 (Promega). The difference between recombinant plasmids pKAR100 and pKAR200 containing a biotinylated peptide and the *argRWT* and *argRNV* genes, respectively is only at two adjacent amino acid residues of the 156 amino acid residues coding sequence for ArgR (Burke *et al.*, 1994; see also Figure 3.15 and Figure 3.16). The two of aspartate residues (D) of

ArgRWT at position 128 and 129 were substituted with asparagine (N) and valine (V) respectively to construct ArgRNV.

In the beginning of this study, I faced some problems in obtaining the recombinant plasmids because the vector and the insert were digested with the same restriction enzyme, *Bam*HI. The insert DNA tends to ligate in the wrong orientation or the vector tends to religate (Figure 3.9). The right recombinant genes may express proteins that are not preferred by the cells.

In vivo cer-mediated recombination assays were carried out as a preliminary study in elucidating the function of the ArgR fusion proteins in Xer site-specific recombination. Intramolecular recombination was measured by the resolution of two reporter plasmids (pCS202 and pSH10) containing dimeric *cer* sites. The results showed that ArgRNV-biotinylated peptide fusion protein supports Xer site-specific recombination *in vivo* better than ArgRWT-biotinylated peptide fusion protein. The plasmid pCS202 was completely resolved to plasmid pCS203 in cells DS956 harbouring pKAR200 expressing ArgRNV-biotinylated peptide fusion protein. ArgRWT-biotinylated peptide fusion protein showed poor Xer site-specific recombination *in vivo* on both reporter plasmids. Several possible consequences of the additional amino acid residues to ArgR structure and activity are discussed below.

6.2 Effects of additional amino acids residues to ArgR structure and activity

The addition of extra amino acid residues (125 amino acids; 13 kDa biotinylated peptide) to the N-terminal domain of ArgR might simply alter the characteristics of ArgR. Previous studies on ArgR domain structure have shown that ArgR subunit is made of two functional regions: a basic N-terminal half residues responsible for DNA binding and an acidic C-terminal residues responsible for oligomerization and for arginine binding (Burke *et al.*, 1994; Tian and Maas, 1994). ArgRWT fusion protein is less proficient in supporting recombination at *cer*. The reduced efficiency of ArgRWT-biotinylated peptide fusion protein to support *in vivo cer*-mediated recombination may be because the protein is less active or alteration of the N-terminal structure of the protein, which is responsible for DNA binding activity. The biotinylated peptide may mask or hinder the DNA binding domain to interact properly with the DNA target site. Since ArgRWT forms hexamers in solution, one might predict that the ArgRWT-biotinylated peptide fusion protein would be too bulky and result in less efficient interaction (direct and indirect) with the DNA target site and/ or with PepA, XerC and XerD, thus disrupting the formation of an active recombination synapse.

Merican (1995) constructed two ArgRWT and ArgRNV fusion proteins: fusion of ArgRWT or ArgRNV to maltose binding protein (MBP-ArgRWT and MBP-ArgRNV) and fusion of ArgRWT or ArgRNV to 6 histidine residues (His_6-ArgRWT and His_6-ArgRNV). The fusion derivatives of ArgRWT and that of ArgRNV were active *in vivo* and proficiently supported *cer*-mediated

recombination *in vivo*. The results presented in Chapter 4 showed that ArgRWT-biotinylated peptide fusion protein behave differently with ArgRNV-biotinylated peptide fusion protein. One possible reason may be due to the size of the fused segment of which the MBP is too big in size (42.7 kDa) while the 6 histidine residues is too small (0.8 kDa) while that of biotinylated peptide is moderate in size (13 kDa).

This study has shown that the ability of ArgRNV-biotinylated peptide fusion protein to support Xer site-specific recombination at *cer* is not eliminated by the addition of 125 residues in the N-terminus of ArgRNV. ArgRNV-biotinylated peptide fusion protein does not adversely affect the DNA binding properties of the protein compared to ArgRWT-biotinylated peptide fusion protein. The fact that ArgRNV-biotinylated peptide fusion protein proficiently support Xer site-specific recombination compared to ArgRWT-biotinylated peptide fusion protein indirectly support the notion by Burke *et al.* (1994) that ArgRNV binds to its DNA target as trimers. In addition to the above explanation on the molecular structure of ArgRWT and ArgRNV, the processes of bacterial expression of ArgR fusion derivatives could conceivably cause loss or reduction of oligomerization or DNA binding activity as a result of altered solubility or protein folding thus affecting Xer site-specific recombination.

6.3 Expression of *argR* fusion genes and the related properties of the fusion products

Many studies of protein structure-function relationship demand the ability to express a cloned gene product in bacterial or eukaryotic cells followed by the purification of the protein in large quantities for further studies. Construction of fusion protein is a very useful strategy for purification and studying proteins in *Escherichia coli* (Makrides, 1996). C-terminal fusion is a commonly used arrangement, in which the gene encoding protein of interest is fused to the C-terminus of a gene which encodes a highly expressed protein or a peptide, which has a high affinity for a particular ligand. The affinity tag of fusion partner is important for the purification of the fusion proteins for any other purposes in protein studies.

An open reading frame (ORF) of each recombinant plasmid was constructed as shown in Figures 3.15 and Figure 3.16 for plasmids pKAR100 and pKAR200, respectively. Protein expression is under the control of the inducible hybrid *trp/lac*(*tac*) promoter, which is useful for regulated expression of cloned genes in *Escherichia coli* (deBoer *et al.*, 1983; Amann *et al.*, 1983). There are several recognition sites for restriction enzymes in the *argR* sequences, either in plasmids pKAR100 or pKAR200 as the original *argR* sequences (Lim *et al.*, 1987) and various recognition sites at purification tag coding sequence. Induction of the *tac* promoter was achieved by the addition of 1 mM IPTG to a mid-log culture followed by further incubation.

A suitable expression system for recombinant proteins in *E. coli* is necessary for the high-level production of the proteins, which depends on several factors (reviewed by Makrides, 1996). These factors include suitable promoter, expression level, cell growth characteristics, and biological activity of the protein of interest. Expression level is determined by the rate of protein synthesis as well as the rate of degradation products. *Escherichia coli* for many reasons is still the most favourable organism for the high-level production of recombinant proteins. However, not every gene can be expressed efficiently in this organism. Possible explanation include degradation of the protein by host cell proteases (Gottesman and Maurizi, 1992; Carter, 1990; Bachmair and Varshavsky, 1989; Cheng *et al.*, 1981; Goldberg and St. John, 1976; Goldberg and Dice, 1974) and the potential toxicity of the protein (Yike *et al.*, 1996).

6.3.1 Protein degradation systems

One major problem in achieving high-level production of recombinant proteins is the existence of selective protein degradation system. In *E. coli*, approximately 1-2% of the total cell proteins is degraded to amino acids per hour (Rogers *et al.*, 1986; Goldberg and Dice, 1974; Goldberg and St. John, 1976). Van Duyne *et al.* (1996) reported that ArgRWT was partially degraded by endogenous *E. coli* proteases.

The cell's proteolytic apparatus recognizes recombinant proteins as abnormal or foreign proteins and will degrade them rapidly. An important function of protein degradation in bacterial cells as well as in other organisms is to

serve as a cellular defence system to eliminate non-functional proteins whose accumulation could be toxic or interfere the host's metabolism (Goldberg and Goff, 1986). In addition, commercial plasmid vectors used for construction of fusion protein usually contains specific cleavage sites which are recognized by specific proteases to release the protein of interest from the fusion protein (Carter, 1990). In this study, the plasmid PinPoint™ Xa-3 Vector contains a specific cleavage recognition site for Factor Xa protease in the linker sequences, that is at position 379-390 with the respective residues, Ile (isoleucine), Glu (glutamic acid), Gly (glycine), and Arg (arginine). Factor Xa protease will cleave after arginine (R).

Fusion proteins may contain additional cleavage sites for endogenous *E. coli* proteases as well as some other determinants, which increase protein instability (Makrides, 1996; Carter, 1990). Other determinants of protein instability include the "N-end rule" formula, which relates to destabilizing of amino terminal residue of the proteins (Bachmair and Varshavsky, 1989). Another determinant of protein instability is a specific lysine residue located near the amino terminus. This residue acts as an acceptor of a multiubiquitin chain that facilitates protein degradation in eukaryotes (Bachmair and Varshavsky, 1989).

6.3.2 Potential toxicity of the protein

Toxicity of certain molecules to the host is not restricted to foreign proteins but may also result from overexpression of certain native proteins (O'Connor and Timmis, 1987) that can effectively kill the cells. Thus, toxicity

can cause a decreasing of cell growth and automatically reduce the production of recombinant proteins as well as plasmid instability (Bentley *et al.*, 1990). The mutant *argR* gene (*argRNV*) on plasmid pAM401 and pKAR200 may produced toxic ArgR protein which may then reduced the plasmid copy number. This may explain my observation that pAM401 and its recombinant plasmid (pKAR200; Figure 3.12 Panel A) have less copy number than the wild-type plasmid pAM204 and its recombinant plasmid (pKAR100; Figure 3.10). A transcriptional regulation system is then necessary to tightly regulate the promoter for the synthesis of the proteins that may be lethal to the host cells (Doherty *et al.*, 1993).

6.4 Use of ArgR fusion protein

This study has shown that elongated fusion derivates of ArgR is active (although ArgRWT-biotinylated peptide fusion protein, may be less active than ArgRNV-biotinylated peptide fusion protein). The fusion proteins can be used to determine the stoichiometry of ArgRWT- and ArgRNV- DNA complexes. Burke *et al.* (1994) showed that ArgRNV appears to bind its DNA target site as a trimer in gel retardation assays in contrast to ArgRWT which binds DNA as hexamer.

There are several methods that can be used to determine the molecular subunit of ArgRWT and ArgRNV that binds to its binding sites. These include: (1) The double-label method (Fried and Crothers, 1983; Garner and Rezvin, 1982; Rye *et al.*, 1993). The double-label method involved differentially labelling the protein and DNA molecules with different isotopes (radioisotope labelling or fluorescent dye labelling) of known specific activities. The molar ratios of protein

and DNA in a given complex are measured by quantitating the relative amounts of the isotopes in the complex band. (2) The double-cross assay (Kato *et al.*, 1992) which involved two cross-linking steps. The double-cross assay provide direct physicochemical evidence for the oligomerization state of DNA-bound protein. In the primary step (protein-DNA cross-linking step), the protein is UV photo-cross-linked to a radiolabelled oligonucleotide probe. The protein-DNA complex is then electrophoresed under denaturing condition. This step allows the determination of the apparent molecular mass of the protein, which directly contact the DNA probe. In the secondary step (protein-protein cross-linking step), glutaraldehyde cross-linking of protein-protein interactions will allow identification of a new species whose apparent molecular mass is a summation of the apparent molecular masses of each of the components of the DNA-protein complex. (3) The mixed oligomer method (Hope and Struhl, 1987). Hope and Struhl described an approach called the mixed oligomer method to determine the subunit structure of GCN4, the *Saccharomyces cerevisiae* transcriptional activator protein. This approach exploits the differences in gel mobility of the wild-type protein with its truncated or elongated derivatives (which exhibit wild-type binding and multimerization). Gel retardation of an equimolar mixture of two size variants of the same protein which band as a monomer to their DNA-target site should generate two retarded complex bands of equal intensities, when the binding stoichiometry is 1:1. A complex containing a dimeric protein bound to a single site should produce three bands in a 1:2:1 ratio (provided that the two different size variants bind to the target with similar affinity and form equally stable homo and hetero-oligomers). The mixed oligomer method does not require

that the protein be labelled. The fusion derivatives of ArgRWT and ArgRNV can be used for such purpose.

6.5 Possible experiments to be carried out

Further studies can be performed in order to exploit the structure-function relationship of ArgR. Such studies include (1) *in vitro cer*-mediated recombination assay; (2) *in vitro* DNA-protein binding analysis; or (3) three dimensional molecular modelling studies to predict the interaction of ArgRWT- and ArgRNV-fusion derivatives with DNA target sites.

6.6 Conclusions and suggestions

Research on ArgR is still progressing through physiological, genetic and biological phases, gradually towards understanding the molecular nature of the repressor and its interaction with DNA targets. ArgR has a specific structure, functioning as a hexamer. It differs from most of the other bacterial repressors, which functions as dimers or tetramers. By its specific nature, ArgR may have roles in other cellular processes, as seen from an observation that ArgR can bind specifically to other DNA fragments that are not involved in either arginine metabolism or plasmid stability (Stirling *et al.*, 1988). If this assumption is true, ArgR can be compared to other proteins involved in the diverse roles in organism such as the *E. coli* integration host factors (IHF). IHF appears to be involved in many cellular processes such as in site-specific recombination, transposition, plasmid replication, conjugal transfer and transcriptional control of biosynthetic and catabolic operons. IHF is also involved in the excision and packaging of

lambda prophage (Kosturko *et al.*, 1989), and also functions as a specific transcription factor (Craig and Nash, 1984). IHF and ArgR thus typify a class of proteins having some of the features of both gene organizers and gene regulators (D'Ari *et al.*, 1993), making the mechanism of their interaction with DNA particularly interesting. Furthermore, by its specific structure functioning as hexamer, the study of ArgR structure will become a focus to scientist on the various functions of ArgR either as a repressor of arginine biosynthesis in *E. coli*, as an accessory factor in Xer site-specific recombination, or other functions in bacterial systems. I believe that the more complex the structure of a protein, its interaction with other proteins and/or with DNA in the formation of a higher order DNA-protein complex will be more impeccable.

The study of ArgR structure-function relationship still cannot explain the connection between ArgR functions in the regulation of L-arginine biosynthesis and in Xer site-specific recombination. Therefore, further studies of structure-function relationship of ArgR must be carried out, in order to understand more about the ArgR structure-function relationship in many biological processes.